Push Method of Chinese Online Education Personalized Course Content for Foreign Students

Jiaxiu Han[1(✉)] and Miao Guo[2]

[1] China West Normal University International College, Nanchong 637000, China
13458214202@163.com
[2] Shanghai Civil Aviation College, Shanghai 200232, China

Abstract. In order to improve the quality of Chinese online education personalized course content push service, the weighted information entropy is introduced to design and research the content push service of Chinese online education. Based on the interest preference of Chinese online learning of foreign students in colleges and universities in China, complete the construction of personalized portrait of foreign students in colleges and universities in China; extract features from the contents of personalized courses of Chinese online education; calculate the similarity between the contents and images of various courses by using weighted information entropy; and realize the push of the contents of personalized courses of Chinese online education of foreign students in colleges and universities in China by using collaborative filtering algorithm. Through comparative experiments, it is proved that this method can improve the classification and push precision of course content, shorten the execution time and recovery time.

Keywords: Information Entropy · College Students Studying in China · Chinese Education · Online Education · Personalization · Online Courses · Content Push

1 Introduction

At present, colleges and universities to study in China exist in a variety of circumstances, the need for Chinese language learning to adapt to China's higher education. Specifically, from the perspective of nationality distribution, students are mainly from Thailand, South Africa, Morocco, Bangladesh, Kazakhstan, South Korea and other countries. Students come from non-English-speaking countries and need to be assisted by a second or even a third language. From the source of students, it is divided into academic students and non-academic students, of which non-academic students accounted for the majority. At the language level, some students had taken short-term courses before they came to China, and could understand the basic classroom language and dialogue. At the age level, degree students are exchange students, while non-degree students include college preparatory students and social workers from all over the world. From the point of view of internal needs, the students with academic qualifications are the auxiliary study of

G. Gui et al. (Eds.): eLEOT 2023, LNICST 543, pp. 126–138, 2024.
https://doi.org/10.1007/978-3-031-51465-4_9

the course of this major for the completion of Chinese teaching; the students without academic qualifications are divided into the students with normal communication and exchange in life, and other students can continue to study the professional contents of the selected major only after they have systematically studied Chinese and passed the HSK examination. According to the curriculum arrangement and standards, the students studying in China not only have the knowledge and content of Chinese, but also have the cross-cultural content, such as the advanced study of Chinese culture courses [1]. The following courses are offered according to the types of teaching and students' needs: HSK training course, Chinese comprehension, Chinese reading and writing, Chinese spoken language, Chinese listening and Chinese culture. Many foreign students cannot come to China to study Chinese, they need to study Chinese and take Chinese examinations in advance. Therefore, we need to pay attention to their online learning.

Make use of the information and network platform to integrate the Internet and Chinese international education, make use of the advantages and characteristics of the network, create new opportunities for the development of Chinese international education, make Chinese international education adapt to the new development. With the increasing demand for Chinese language learning all over the world, how to provide effective Chinese language learning resources and Chinese language learning platforms for foreign students in colleges and universities will become an important task for the development of international Chinese language education. The construction of online course content resources is an important basis for promoting the development of Chinese language learning of foreign students in colleges and universities.

The motivation for researching personalized course content push methods for Chinese online education for international students is to meet their personalized learning needs, improve learning effectiveness, increase learning motivation and interest, and provide diverse learning resources. By customizing course content and adjusting it according to students' learning progress and ability level, it can better help international students understand and master Chinese language knowledge. At the same time, various types of textbooks, learning materials, and multimedia resources are provided to enrich students' learning experience. This personalized course content push method can enable international students to actively participate in learning, improve the effectiveness and efficiency of learning. In recent years, a large number of Chinese learning platforms and Chinese learning resources have been emerging, which has brought trouble to the use of Chinese learning resources by college students. Under this background, it is extremely urgent to carry on the push research to the on-line Chinese teaching curriculum. A scholar [4] has proposed a trust awareness recommendation approach based on a deep sparse automatic encoder to generate the user's potential characteristics, using reliability metrics to select implicit ratings to enhance rating profiles for users with insufficient ratings. Reliable enhanced rating profiles and trust statements can be used as input to deep sparse autoencoders to generate the user's potential characteristics. Calculates similarity between users and makes recommendations. One scholar [5] created an initial glossary of terms from the existing user learner profile and used the course ontology to determine the exact needs of users. Metadata is initially classified using the XGBoost algorithm, and core data sets are classified using rich user terms. Then, the concept similarity is calculated by genetic algorithm to calculate semantic similarity, which is recommended to users. In practical application, all kinds of push methods do not need the user to provide a clear goal, but can directly find the information that the user may be interested in. However, there are still some drawbacks in the application of push method. For example, the resources pushed for users are too similar to the content of user history query records. Based on the introduction of weighted information entropy, this paper tries to design and study the individualized push method of Chinese online education curriculum resources. The innovation of this research lies in the use of weighted information entropy to calculate the similarity between course content and image, and the combination of Collaborative filtering algorithm to achieve the construction of personalized image of foreign students in Chinese universities and the push of online education course content, so as to meet their interest in online learning in China.

2 Personalized Portrait Construction of Chinese Online Education for Foreign Students in Colleges and Universities

In the online Chinese education for college students studying in China, the basic information of college students studying in China and the data that can indicate the users' interests and preferences shall be obtained, and these information and data shall be cleaned and converted to generate labels through statistics and analysis, and the construction of personalized portraits of the users of the online Chinese education platform for college students studying in China shall be completed [6]. In order to improve the accuracy of the push method, a personalized portrait of Chinese online education for college students in China is constructed by integrating the interests of college students in China, and the interest of college students in China is represented by a spatial model:

$$U = \{(T_1, W_1); (T_2, W_2); \cdots ; (T_n, W_n)\} \tag{1}$$

In the formula, U represents the spatial model representation of the user interest of the university students studying in China; T_1, T_2, ..., T_n represents the dimension label of the user interest of the university students studying in China; W_1, W_2, ..., W_n represents the weight of a dimension label of interest. When creating the personalized portrait of Chinese online education for foreign students in colleges and universities, the corresponding portrait is generated according to the classification number of the contents of the education courses read by foreign students in colleges and universities. Assuming that the number of occurrence of the classification number at the end of the users of the returned university students in China is greater than or equal to m, the label in the portrait shall be "classification number", and the weight shall be "the number of times the corresponding course content of the classification number has been borrowed by the returned university students in China", and m shall be the subsequent filter parameter condition. In addition, in the content of online Chinese education courses, there are also user message boards, comment data and other resource information of some overseas students of colleges and universities in China. For this part of content, the method of manual labeling may be adopted [7] to take the key words in the resource information as the portraits of overseas students of colleges and universities in China, and continue to assign weight to the personalized portraits of overseas students of colleges and universities in China in accordance with the specific content of messages and emotions contained in comments of overseas students of colleges and universities in China [8]. Based on the above contents, we can not only describe the users of Chinese online education, but also reduce the operation burden of Chinese online education platform and provide favorable conditions for the subsequent courses.

3 Feature Extraction of Chinese Online Education Courses

Firstly, it is necessary to determine the distribution of the content data of online Chinese education courses. In the environment of heterogeneous data storage of the content data of online Chinese education courses, the preference for decision-making characteristics of the content data of online Chinese education courses shall be introduced, and the integration of the personalized portrait of users of overseas Chinese students of colleges and universities in China and the content characteristics data of online Chinese education courses shall be realized [9]. In this process, the following formulas may be used:

$$Q(x) = \frac{P(x)}{\sum_{i=1} (V_i(x) - P_i(x))} \tag{2}$$

In the formula, $Q(x)$ represents the characteristic data of the content of online Chinese education courses obtained through the above operation; $P(x)$ represents the personalized portrait label of the users of overseas Chinese students of colleges and universities corresponding to the characteristic data; $V_i(x)$ represents the characteristic value of all the content of online Chinese education courses; $P_i(x)$ represents the personalized portrait label of the users of overseas Chinese students of colleges and universities corresponding to a specific content of online Chinese education courses. On the basis of the above formula, the integration of the personalized portraits of foreign students in colleges and universities and the feature data of the content of online Chinese education courses needs to be completed, and on the basis of this, the integration of the feature data of the content of online Chinese education courses itself is also required. The process can be expressed by the following formula:

$$S = \varphi^s \cdot \delta + \varphi^b \cdot A \tag{3}$$

In the formula, S represents the objective function for feature extraction of the content of a Chinese online education course; φ^δ represents the weight coefficient for feature extraction under δ push delay condition [10]; and φ^A represents the feature extraction weight when the consumption is A in the feature extraction of a certain Chinese online education course content. Based on the above operations, the integration of various resources in the Chinese online education curriculum should be realized. After the integration, the characteristic data should be reconstructed in multi-dimensions [11]. The data can be reconstructed by the method of fusion and matching, and the reconstructed data can be imported into the Chinese online education course content push database.

4 Course Content Similarity Calculation Based on Weighted Information Entropy

Weighted information entropy is a measure of data uncertainty. It is the information entropy of weighted data. In information theory, entropy is a measure of uncertainty of random events. It represents the average amount of information contained in all events in a set of events. Weighted information entropy takes into account the weight of each event when calculating entropy, which can be any positive number [12]. Weighted information entropy can be used in data compression, data classification, data analysis and other fields. The implementation steps are shown in Fig. 1 below:

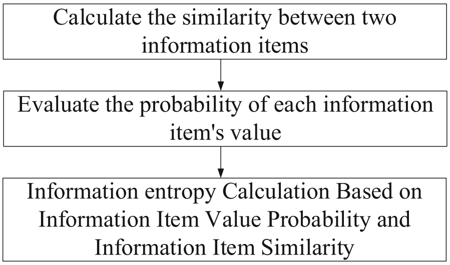

Fig. 1. Flow of information entropy calculation

In order to improve the efficiency of pushing and reduce the burden of pushing, weighted information entropy [13] is introduced to calculate the similarity of Chinese online education courses.

During the calculation, the information entropy is used to describe the push probability of the course content:

$$\begin{bmatrix} X \\ p \end{bmatrix} = \begin{bmatrix} X_1 & X_2 & \cdots & X_n \\ p_1 & p_2 & \cdots & p_n \end{bmatrix} \tag{4}$$

In the formula, X represents a random variable for discrete course content; p represents X push probability distribution. The lower the information entropy is, the more orderly the curriculum content is, and the higher the information entropy is, the more chaotic the curriculum content is. Assuming that the information entropy of a certain course content in the Chinese online education course can be expressed as $H(X)$, the following formula may be used as the basis for calculation:

$$H(X) = r[y(x_i)] \tag{5}$$

In the formula, $y(x_i)$ represents the logarithmic negative value of the probability of a given course content being pushed; r represents the result of the ordered degree measurement of Chinese online education content [14]. Convert formula (5) further to the following formula:

$$H(X) = \sum_{i=1}^{n} p_i \log_2 \frac{1}{p_i} \tag{6}$$

In the formula, p_i represents the probability that category I of an online Chinese course is pushed. The information entropy can be used to describe the confusion degree of the content information of online Chinese education courses, and to provide a basis for judging the unity of the content information. When there is only one category in the content of a Chinese online education course, the value of p_i is 1 and the value of $H(X)$ is 0. When the probability of pushing n classifications is the same, then p_i turns to the reciprocal of the number of classifications, that is 1/n, $H(X)$ has the maximum value, $\log_2 n$. After the content similarity of Chinese online education courses is calculated, the basis for subsequent collaborative filtering of the course content is provided. On this basis, we can also measure the similarity between the content of Chinese online education courses and the interest of foreign students in China. Assuming a grade difference of D between one user A and another user B for the same course content, the following formula can be used to calculate the value of g:

$$g = |g_A - g_B| = (|d_1|, |d_2|, \cdots |d_n|) \tag{7}$$

In the formula, g_A means that user A scores the content of Chinese online education courses; g_B means that user B scores the content of Chinese online education courses; $d_1, d_2, \ldots, d_n$ means that user A scores the difference between user B and user A on each item of Chinese online education courses. After calculating the difference between user A and user B on the content of the same Chinese online education course, this difference is analyzed in frequency. Classify the difference value and calculate the probability of each classification. Assuming that in the process of frequency statistics, g is subdivided into M values, in extreme cases, when each value of g is different, M = N exists; when each value of g is the same, M = 1 exists. The total frequency of g occurrence can be obtained by summarizing the frequency of the occurrence of a category. The weighted information entropy of g is calculated in combination with the above formula (6). The range of information entropy is 0–+∞. When the value of information entropy is 0, the similarity between user A and user B is very similar. By using the above methods, we can not only calculate the similarity of two universities' students studying in China, but also calculate the similarity of many universities' students studying in China. For example, suppose that between user A and user B, the g value of user A is 1, and the g value of user B is 2; between user A and user C, the g value of user A is 3, and the g value of user C is 4, and if the similarity between user A and user B is 1, then the similarity between user A and user C is also 1. Based on the above logic, we determine the similarity between all the users of Chinese online education courses, and carry out subsequent collaborative filtering and push of the courses.

5 Collaborative Filtering Push of Chinese Online Education Course Content Resources

Through the above calculation, we can get the similarities between the contents of Chinese online education courses and those of foreign students in China, and feedback the preferences of foreign students in China. For several users of Chinese online education courses, a user rating matrix can be established according to the way shown in Fig. 2.

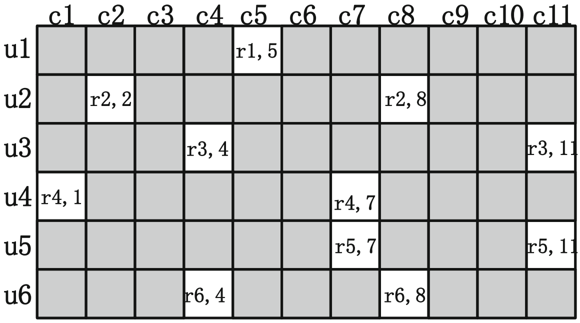

Fig. 2. Scoring matrix of university students studying in China for course contents

In Fig. 2, u represents foreign students from universities in China, c represents the content of the course, and rij represents the results given by users of foreign students in universities in China. The gray areas in the image indicate that university students coming to China did not receive or grade the course content. In Fig. 1, RIJ can be divided into two types: explicit feedback and implicit feedback according to the specific forms of Chinese students' behavior. Based on the feedback results given by overseas students from colleges and universities in China, the following formula is used as the threshold for collaborative filtering of resources:

$$f = r_A + \frac{Sim(A, B) * (r_A - r_B)}{\sum Sim(A, B)} \tag{8}$$

In the formula, f represents the collaborative filtering threshold for resources; r_A represents the scoring results given by user A for all the evaluated push course content; $Sim(A, B)$ represents the similarity of scoring results given by user A and user B for the same push course content; and r_B represents the scoring results given by user B for all the evaluated push course content. Take the value of f as the basis for collaborative filtering of resources. If the threshold of collaborative filtering of resources for a given course content is calculated and the result is greater than f after calculation, the course content does not meet the push conditions and is filtered from the push results[15]; if the result is less than or equal to f, the course content meets the push conditions and is retained in the push results and displayed on the real interface for overseas students in China, so as to achieve the individualized course content of online Chinese education for overseas students in China (Fig. 3).

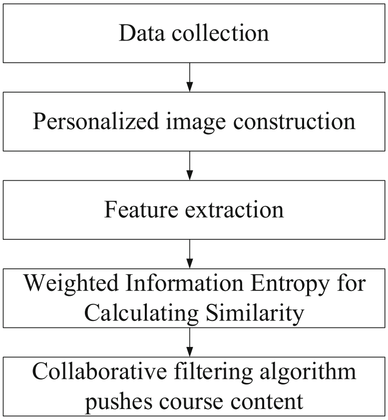

Fig. 3. Flow chart of the algorithm implementation

6 Experimental Results and Analysis

6.1 Experimental Preparation

But this method is still at the stage of theoretical research. In order to popularize this method and evaluate its effect, we need to design and compare experiments on the basis of existing research. Therefore, the following will take the design of comparative experiments to test the effectiveness of this method. Using literature [4] method and literature [5] method as the control group, this paper compares and analyzes the recommendation effects of three recommendation methods in the same experimental environment.

In order to test the effectiveness of the content push method of Chinese online education personalized course for foreign students in colleges and universities in China, a simulation experiment is conducted (Table 1).

Table 1. Test environment

Environmental parameter	Parameter value
CPU	Intel Cy Young G4900
RAM	Kingston DDR4 2400 16G
OS	Windows 10
Programming languages	Java
Programming environment	Eclipse3.2

The resources used in the experiment are a university database, which extracts Chinese online education courses and other related contents for foreign students in China as data sets. The dataset contains 20, 000 pieces of data and is divided into five datasets, as shown in Table 2.

Table 2. Data sets tested

Data set number	Sample size
1	2000
2	1000
3	3000
4	1500
5	2500

6.2 Comparison of AUC Indicators

The AUC value is chosen as the evaluation index. The AUC value refers to the surrounding area under the work characteristic curve ROC. The higher the AUC value is, the better the effect of the recommendation method is, the more reasonable the subsequent recommendation is. In the course of the experiment, three recommendation methods are set up to classify the personalized course content of Chinese online education for foreign students in colleges and universities in China with the data volume of 200 Mbits to 1000 Mbits, and in the course of classification, the ROC values of each of them are calculated according to the following formulas:

$$ROC = (l - L)/L \tag{9}$$

In the formula: l means the data of the correct classification of the personalized course content of online Chinese education for overseas Chinese students in colleges and universities; L means the total data of the personalized course content of online Chinese education for overseas Chinese students in colleges and universities. The ROC values of the three recommended methods are calculated according to the above formulas, and the results are drawn as the experimental results shown in Fig. 4.

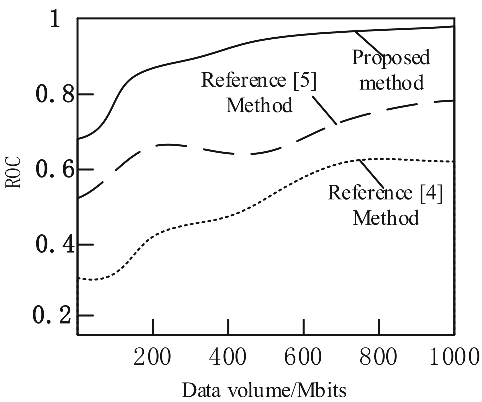

Fig. 4. Comparison of AUC values

As can be seen from the two curves in Fig. 4, the ROC value of this method is obviously higher than that of Reference [4] Method and Reference [5] Method. Therefore, through the above experimental results, it is proved that the proposed method can classify the content of personalized course of Chinese online education for foreign students in China accurately in practical application, and provide more favorable conditions for pushing the content of personalized course of Chinese online education for foreign students in China.

6.3 Comparison of Push Accuracy

Based on the comparison of the three methods, 5 international students were selected as the experimental volunteers, who were numbered A, B, C, D and E. Three push methods are used to push the corresponding Chinese online education personalized course content according to the students' preference. Compared with the click rate, the push precision is quantified. The results are recorded in Table 3.

Table 3. Comparison of click-through rates for push course content

User	Click through rate of this method (times/100 Mbits)	Reference [4] Method click rate (times/100 Mbits)	Reference [5] Method click-through rate (times/100 Mbits)
User A	23.65	9.22	15.52
User B	24.72	8.54	16.34
User C	25.86	6.26	12.76
User D	27.24	7.25	14.36
User E	29.55	9.62	15.33

From the experimental results in Table 3, it can be seen that the click-through rate of the proposed method is more than 22 times, the click-through rate of the reference [4] method is less than 10 times, and the reference [5] method is less than 17 times. It can be seen that the Chinese online education personalized course content pushed by this method is more welcomed by the overseas Chinese students, and more in line with the actual needs of the Chinese online education personalized course content.

6.4 Implementation Time Comparison

The average execution time of this method is 238 ms, the average execution time of reference [4] method is 467 ms, and the average execution time of reference [5] method is 512 ms. Compared with other methods, the implementation time of this method is significantly reduced, and the speed of content push is accelerated. The efficiency of this method is higher (Fig. 5).

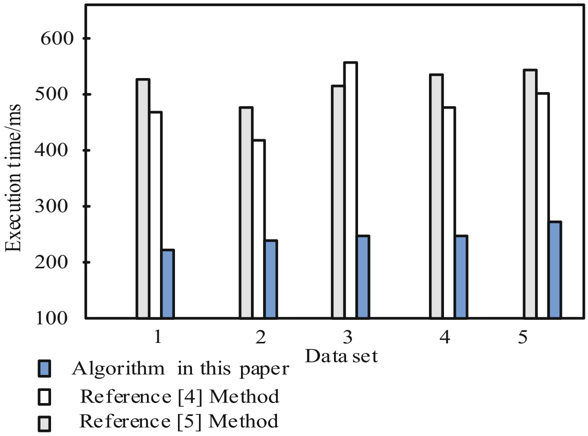

Fig. 5. Execution Time Comparison/ms

6.5 Reliability Comparison

In order to test the reliability of the method, some unstable conditions are artificially created. When the method changes in the working environment of the platform, the result is shown in Table 4. It can be found from Table 4 that the method of pushing the individualized course content of Chinese online education for foreign students in colleges and universities in China has reached a stable state in a short time, with the scope of [100.2, 184.2], while the time for the method of reference [4] method and reference [5] method to reach a stable state is obviously prolonged, with the scope of [105.6, 390.9] and [319.5, 405.6], respectively, which indicates that the method of this paper is more reliable, more able to adapt to the change of external environment, and has stronger robustness.

Table 4. Reliability Comparison

Degree of change in working environment	This method achieves steady state time/ms	Reference [4] method to reach steady state time/ms	Reference [5] method to reach steady state time/ms
Very large	184.2	390.9	405.6
Big	167.7	207.1	388.6
General	130.6	189.8	341.5
Small	114.7	158.2	327.6
Very small	100.2	105.6	319.5

The method proposed in this paper stands out from traditional approaches in several aspects. Firstly, it demonstrates superior application performance by providing different results tailored to different types of overseas Chinese students. The classification of course content based on personalized interests achieves higher ROC and AUC values, enabling the delivery of courses that align with the specific preferences of individual overseas Chinese students.

Furthermore, the verified method in this paper successfully pushes personalized course content for online Chinese education to overseas Chinese students over 22 times. This indicates a significant level of interest among university-level overseas Chinese students in the content recommended by the proposed method, surpassing the outcomes of alternative approaches. The average time required for content push in this method is 238, which is remarkably fast. This not only enhances user satisfaction but also facilitates regular online education activities for overseas Chinese students in universities.

It is worth noting that the method in this paper exhibits greater stability compared to other approaches when subjected to external disturbances. The recovery time from disturbance to stability is significantly shorter at 18.24 ms. This reliability contributes to the overall effectiveness and comprehensive promotion of the proposed method.

In conclusion, the distinctive features of this method, including its ability to deliver personalized course content and its superior stability and reliability, make it a valuable technical support for online Chinese education among college-level overseas Chinese students.

7 Closing Remarks

Based on the introduction of weighted information entropy, this paper puts forward a method of pushing the content of Chinese online education personalized course for foreign students. Based on the analysis of user portraits of foreign students in Chinese colleges and universities, this paper extracts the features of personalized course contents of Chinese online education. Through experiments, the advantages of the push method are verified from the classification accuracy and push accuracy of the course content. From the analysis of execution time and reliability index, the performance of push method is verified. In the practical application, if we can push the content of Chinese online education for the students in China according to the above ideas, we can not only provide the service, but also promote the overall operation quality of the platform. In future research work, we will build an online learning platform to provide an interactive and social learning environment for international students, encourage them to communicate, share experiences and resources with each other, and enhance the interactivity and sociality of learning.

References

1. Ma, G., Black, K., Blenkinsopp, J., et al.: Higher education under threat: China, Malaysia, and the UK respond to the COVID-19 pandemic. Compare J. Comp. Int. Educ. **52**(5), 841–857 (2022)
2. Manegre, M., Sabiri, K.A.: Online language learning using virtual classrooms: an analysis of teacher perceptions. Comput. Assist. Lang. Learn. Assist. Lang. Learn. **35**(5–6), 973–988 (2022)

3. Agyeiwaah, E., Baiden, F.B., Gamor, E., et al.: Determining the attributes that influence students' online learning satisfaction during COVID-19 pandemic. J. Hosp. Leis. Sport Tour. Educ. **30**, 100364 (2022)
4. Ahmadian, M., Ahmadi, M., Ahmadian, S.: A reliable deep representation learning to improve trust-aware recommendation systems. Exp. Syst. Appl. **197**, 116697 (2022)
5. Agrawal, D., Deepak, G.: HSIL: hybrid semantic infused learning approach for course recommendation. In: Motahhir, S., Bossoufi, B. (eds.) Digital Technologies and Applications, ICDTA 2022. LNNS, vol. 454, pp. 417–426. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-01942-5_42
6. Toro-Domínguez, D., Martorell-Marugán, J., Martinez-Bueno, M., et al.: Scoring personalized molecular portraits identify Systemic Lupus Erythematosus subtypes and predict individualized drug responses, symptomatology and disease progression. Brief. Bioinform. **23**(5), bbac332 (2022)
7. Kuang, A.: Construction of personalized advertising accuracy model based on artificial intelligence. In: 2022 International Conference on Artificial Intelligence and Autonomous Robot Systems (AIARS), pp. 395–398. IEEE (2022)
8. Shakhovska, N., Fedushko, S., Melnykova, N., et al.: Big Data analysis in development of personalized medical system. Procedia Comput. Sci. **160**, 229–234 (2019)
9. Buchaiah, S., Shakya, P.: Bearing fault diagnosis and prognosis using data fusion based feature extraction and feature selection. Measurement **188**, 110506 (2022)
10. Calvini, R., Pigani, L.: Toward the development of combined artificial sensing systems for food quality evaluation: a review on the application of data fusion of electronic noses, electronic tongues and electronic eyes. Sensors **22**(2), 577 (2022)
11. Patil, R.R., Kumar, S.: Rice-fusion: a multimodality data fusion framework for rice disease diagnosis. IEEE Access **10**, 5207–5222 (2022)
12. Sidhu, A.S., Singh, S., Kumar, R.: Bibliometric analysis of entropy weights method for multi-objective optimization in machining operations. Mater. Today Proc. **50**, 1248–1255 (2022)
13. Siddique, M.A.B., Islam, A.R.M.T., Hossain, M.S., et al.: Multivariate statistics and entropy theory for irrigation water quality and entropy-weighted index development in a subtropical urban river, Bangladesh. Environ. Sci. Pollut. Res.Pollut. Res. **29**, 8577–8596 (2022)
14. Aydoğdu, A., Gül, S.: New entropy propositions for interval-valued spherical fuzzy sets and their usage in an extension of ARAS (ARAS-IVSFS). Exp. Syst. **39**(4), e12898 (2022)
15. Li, H., Jian, W.: User big data cycle intelligent recommendation algorithm based on collaborative filtering. Comput. Simul. **40**(03), 476–479+489 (2023)